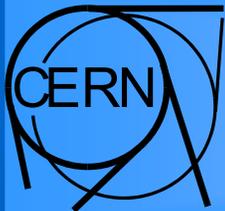


Second Large-Scale Cluster Computing Workshop

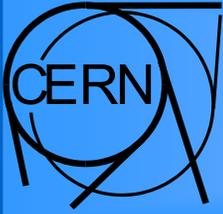
held at Fermilab
21-22nd October 2002

Alan Silverman and Mark Kaletka



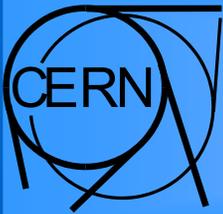
Outline

- Background and Goals
- The first in the series
- Format of this Workshop
- Proceedings
- References



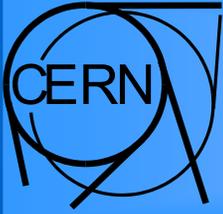
Background

- Sponsored by HEPIX, in particular by the Large Cluster SIG
- In background reading on Grid technologies, we found many papers and USENIX-type talks on cluster techniques, methods and tools.
- But often with results and conclusions based on small numbers of nodes.
- What is the “real world” doing?
- ***Gathering practical experience is the primary goal***



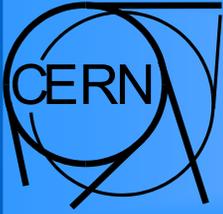
Goals

- Understand what exists and what might scale to large clusters (1000-5000 nodes and up).
- And by implication, predict what might not scale
- Produce *the* definitive guide to building and running a cluster - how to choose/select/test the hardware; software installation and upgrade tools; performance mgmt, logging, accounting, alarms, security, etc, etc
- Maintain this.



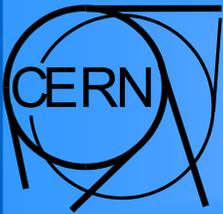
The First Workshop

- Participation was targeted at sites with a minimum cluster size (100-200 nodes)
- Invitations were sent, not only to HENP sites but to other sciences, including biophysics. We also invited participation by technical representatives from commercial firms (sales people refused!)



Workshop Layout

- Apart from a few plenary sessions, typically to set the scale of the problem as compared to where we are today, the workshop was arranged in 2 streams of highly-interactive panels
- Each panel was presented with some initial questions to consider as a starting point
- Each panel was “seeded” with 2 or 3 short informal talks relevant to the panel topic
- The panels were summarised on the last day
- Proceedings were published and reported at CHEP



Cluster Builders Guide

- A framework covering all (we hope) aspects of designing, configuring, acquiring, building, installing, administering, monitoring, upgrading a cluster.
- Not the only way to do it but it should make cluster owners think of the correct questions to ask and hopefully where to start looking for answers.
- Section headings to be filled in as we gain experience.

- **1. Cluster Design Considerations**

- 1.1 What are characteristics of the computational problems ?**

- **1.1.1 Is there a “natural” unit of work ?**

- **1.1.1.1 Executable size**
 - **1.1.1.2 Input data size**
 - **.....**

- 1.2 What are characteristics of the budget available ?**

- **1.2.1 What initial investment is available ?**

- **1.2.2 What is the annual budget available ?**

- **.....**

- **.....**

- **5. Operations**

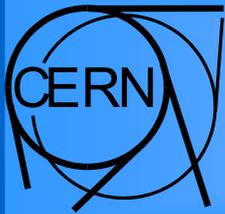
- 5.1 Usage**

- 5.2 Management**

- **5.2.1 Installation**

- **5.2.2 Testing**

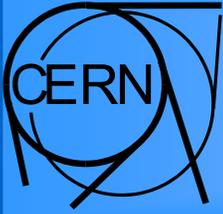
- **.....**



This Meeting

- Two themes, practical experience (again) and the various technologies needed to build a cluster
- It's a workshop, it relies on active participation. Ask questions, make points, discuss.
- There will be as full a summary as I can make but I need help
 - Slides from speakers
 - Notes you take

(alan.silverman@cern.ch)



References

Most of the overheads presented at the workshop should be found on the web site

<http://conferences.fnal.gov/lccws/>

You will also find there the Proceedings (no promise when) and some useful cluster links (including many links within the Proceedings).

Other useful links for clusters

IEEE Cluster Task Force <http://www.ieeetfcc.org>

Top500 Clusters <http://clusters.top500.org>